

# IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

## IN THE CLAIMS

Please amend the claims as follows:

1. (Previously Presented) A method for managed object replication and delivery in a system comprising a network having one or more edge server sites and one or more parent server sites distinct from said edge server sites, the method comprising:

(A) directing a request by a client for an object to a particular edge server site in the network; and

(B) if the particular edge server site does not have the requested object, then

(b1) said particular edge server site redirecting the client request to a first server site distinct from the edge server site; and

(b2) conditionally replicating a portion of the requested object on the particular edge server site for use in serving future client requests, said replicating being based at least in part on a dynamic measure of popularity of the requested object, wherein the portion of the requested object is replicated on the particular edge server site when the dynamic measure of popularity of the requested object exceeds a dynamic replication threshold, said dynamic replication threshold being based at least in part on at least one dynamic measure of capacity, and wherein the requested object is served to the client from a server site other than the particular edge server site.

2. (Previously Presented) The method of claim 1, wherein redirecting the client request to said first server site comprises said particular edge server site redirecting the client request to a parent server site in the network.

3. (Previously Presented) The method of claim 2, wherein if that parent server site does not have the requested object, then recursively redirecting the request until a parent server site in the network having the requested object is reached, and then serving the requested object to the client from the parent server site that has the requested object.

4. (Canceled)

5. (Previously Presented) The method of claim 1, wherein directing a request by a client for an object to a particular edge server site comprises directing the request by the client for an object to a best or optimal edge server site.

6. (Previously Presented) The method of claim 5, wherein a best or optimal edge server site comprises an edge server site selected using at least one of a determination based on a best repeater selector, the likelihood of a copy of the requested object being available at the edge server site, and the bandwidth between the edge server site and the client.

7. (Previously Presented) The method of claim 1, wherein said step of conditionally replicating the portion of the requested object to the particular edge server site comprises replicating the portion of the requested object to the particular edge server site from a parent server site.

8. (Previously Presented) The method of claim 1, wherein said step of conditionally replicating comprises:

if the requested object is determined to be popular based on said dynamic measure of popularity, and if the requested object is unavailable on parent server sites in the network, then replicating the portion of the requested object to a parent server site in the network from an origin server site.

9. (Previously Presented) The method of claim 1, wherein said step of conditionally replicating the portion of the requested object to the particular edge server

site comprises, if the requested object is unavailable on parent server sites in the network, replicating the portion of the requested object to the particular edge server site from an origin server site.

10. (Previously Presented) The method of claim 1, wherein said dynamic measure of popularity of the requested object is determined using at least a request rate for the requested object.

11. (Previously Presented) The method of claim 1, further comprising:

if an object on the particular edge server site is no longer popular, as determined based on said dynamic measure of popularity of said object, deleting at least some part of the object from the particular edge server site.

12. (Previously Presented) The method of claim 1, further comprising:

if an object on the parent server site is no longer popular, as determined based on said dynamic measure of popularity of said object, and if the object is available on an origin server site, then deleting at least some part of the object from the parent server site.

13. (Canceled).

14. (Previously Presented) The method of claim 1, wherein said step of conditionally replicating the portion of the requested object on said particular edge server site comprises:

replicating the portion of the requested object when said dynamic measure of popularity of the requested object is greater than a the dynamic replication threshold and

there is enough storage on said particular edge server site to replicate the portion of the requested object;

otherwise, if there is not enough storage on said particular edge server site to replicate the portion of the requested object, then

i) comparing the dynamic measure of popularity of the requested object against a dynamic measure of popularity of a least popular object in storage on said particular edge server site,

ii) if the dynamic measure of popularity of the requested object exceeds the dynamic measure of popularity of the least popular object in the storage, deleting at least some part of the least popular object from the storage, and then

iii) repeating i) and ii) until enough storage is available for the portion of the requested object or until the dynamic measure of popularity of the requested object is less than the dynamic measure of popularity of the least popular object in the storage, and then

iv) replicating the portion of the requested object on said particular edge server site if there is enough storage on said particular edge server site.

15. (Previously Presented) The method of claim 1, wherein the step of serving the requested object is performed separately from the step of conditionally replicating the portion of the requested object.

16. (Previously Presented) A method for managed object replication and delivery in a system comprising a network having one or more parent server sites and one or more edge server sites distinct from said parent server sites, the method comprising:

(A) directing a request by a client for an object to an optimal edge server site in the network; and

(B) if the edge server site does not have the requested object, then the edge server site redirecting the client request to a parent server site in the network distinct from the edge server site; and,

(C) if the edge server site does not have the requested object, conditionally replicating a portion of the requested object to the edge server site from a parent server site in the network, said replicating being based at least in part on a dynamic measure of popularity of the requested object relative to a dynamic replication threshold, said dynamic replication threshold being based, at least in part, on at least one dynamic measure of capacity,

wherein the requested object is served to the client from a server site other than the edge server site.

17. (Previously Presented) The method of claim 16, further comprising:

if an object on the edge server site is no longer popular, as determined based on said dynamic measure of popularity of said object, deleting at least some part of the object from the edge server site.

18. (Previously Presented) The method of claim 16, further comprising:

if an object on the parent server site is no longer popular, as determined based on the said dynamic measure of popularity of said object, and if the object is available on an origin server site, deleting at least some part of the object from the parent server site.

19. (Canceled)

20. (Previously Presented) The method of claim 16, wherein said step of conditionally replicating the portion of the requested object on said edge server site comprises:

replicating the portion of the requested object when said dynamic measure of popularity of the requested object is greater than a the dynamic replication threshold and there is enough storage on said edge server site to replicate the portion of the requested object;

otherwise, if there is not enough storage on said edge server site to replicate the portion of the requested object, then

i) comparing the popularity of the requested object, as determined based on said dynamic measure of popularity, against a dynamic measure of popularity of a least popular object in storage on the edge server site,

ii) if the dynamic measure of popularity of the requested object exceeds the dynamic measure of popularity of the least popular object in the storage, deleting at least some part of the least popular object from the storage, and

iii) repeating i) and ii) until enough storage is available for the portion of the requested object or until the dynamic measure of popularity of the requested object is less than the dynamic measure of popularity of the least popular object in the storage, and then

iv) replicating the portion of the requested object on the edge server if there is enough storage.

21. (Original) The method of claim 16, wherein whether the requested object is popular is determined using at least a request rate for the requested object.

22. (Previously Presented) The method of claim 16, wherein the step of serving the requested object is performed separately from the step of conditionally replicating a portion of the requested object.

23. (Previously Presented) A computer program product embodied on non-transient computer-readable storage media including computer program code to cause a processor to perform a method for managed object replication and delivery in a system comprising a network having one or more edge server sites and one or more parent server sites distinct from said edge server sites, the computer-readable media comprising computer program code for:

(A) directing a request by a client for an object to a particular edge server site in the network; and

(B) if the particular edge server site does not have the requested object, then

(b1) the particular edge server site redirecting the client request to a first server site distinct from the particular edge server site; and,

(b2) conditionally replicating a portion of the requested object on the particular edge server site for use in serving future client requests, said replicating being based at least in part on a dynamic measure of popularity of the requested object, wherein the portion of the requested object is replicated on the particular edge server site when the dynamic measure of popularity of the requested object exceeds a dynamic replication threshold, said dynamic replication threshold being based at least in part on at least one dynamic measure of capacity, and wherein the requested object is served to the client from a server site other than the particular edge server site.

24. (Previously Presented) The computer program product of claim 23, wherein redirecting the client request to a first server site comprises redirecting the client request to a parent server site in the network.

25. (Previously Presented) The computer program product of claim 23, wherein redirecting the client request to a first server site comprises redirecting the client request to a parent server site in the network, and if that parent server site does not have the

requested object, then recursively redirecting the request until a parent server site in the network having the requested object is reached and then serving the requested object to the client from the parent server site that has the requested object.

26. (Canceled)

27. (Previously Presented) The computer program product of claim 23, wherein directing a request by a client for an object to a particular edge server site comprises directing the request by the client for an object to a best or optimal edge server site.

28. (Previously Presented) The computer program product of claim 27, wherein a best or optimal edge server site comprises an edge server site selected using at least one of a determination based on a best repeater selector, the likelihood of a copy of the requested object being available at the edge server site, and the bandwidth between the edge server site and the client.

29. (Previously Presented) The computer program product of claim 23, wherein the conditionally replicating the portion of the requested object to the particular edge server site comprises replicating the portion of the requested object to the particular edge server site from a parent server site.

30. (Previously Presented) The computer program product of claim 23, the method further comprising:

if the requested object is determined to be popular based on said dynamic measure of popularity, and if the requested object is unavailable on parent server sites in the network, then replicating the portion of the requested object to a parent server site in the network from an origin server site.



31. (Previously Presented) The computer program product of claim 23, wherein said conditionally replicating the portion of the requested object to the particular edge server site comprises, if the requested object is unavailable on parent server sites in the network, replicating the portion of the requested object to the particular edge server site from an origin server site.

32. (Previously Presented) The computer program product of claim 23, wherein said dynamic measure of popularity of the requested object is determined using at least a request rate for the requested object.

33. (Previously Presented) The computer program product of claim 23, the method further comprising:

if an object on the particular edge server site is no longer popular, as determined based on said dynamic measure of popularity of said object, deleting at least some part of the object from the particular edge server site.

34. (Previously Presented) The computer program product of claim 23, the method further comprising:

if an object on the parent server site is no longer popular, as determined based on said dynamic measure of popularity of said object, and if the object is available on an origin server site, then deleting at least some part of the object from the parent server site.

35. (Canceled)

36. (Previously Presented) The computer program product of claim 23, wherein said conditionally replicating the portion of the requested object on said particular edge server site comprises:

replicating the portion of the requested object when said dynamic measure of popularity of the requested object is greater than a the dynamic replication threshold and there is enough storage on said particular edge server site to replicate the portion of the requested object;

otherwise, if there is not enough storage on said particular edge server site to replicate the portion of the requested object,

i) comparing the dynamic measure of popularity of the requested object against a dynamic measure of popularity of a least popular object in storage on said particular edge server site,

ii) if the dynamic measure of popularity of the requested object exceeds the dynamic measure of popularity of the least popular object in the storage, deleting at least some part of the least popular object from the storage, and

iii) repeating i) and ii) until enough storage is available for the portion of the requested object or until the dynamic measure of popularity of the requested object is less than the dynamic measure of popularity of the least popular object in the storage, and then

iv) replicating the portion of the requested object on said particular edge server site if there is enough storage.

37. (Previously Presented) The computer program product of claim 23, wherein serving the requested object is performed separately from conditionally replicating the portion of the requested object.

38. (Previously Presented) A computer program product embodied on non-transient computer-readable storage media including computer program code having instructions to cause a processor to perform a method for managed object replication and delivery in a system comprising a network having one or more parent server sites and one or more edge server sites distinct from said parent server sites, the method comprising:

(A) directing a request by a client for an object to an optimal edge server site in the network; and

(B) if the optimal edge server site does not have the requested object, the edge server site redirecting the client request to a parent server site in the network distinct from said edge server site; and,

(C) if the edge server site does not have the requested object, conditionally replicating a portion of the requested object to the edge server site from a parent server site in the network, said replicating being based at least in part on a dynamic measure of popularity of the requested object relative to a dynamic replication threshold, said dynamic replication threshold being based at least in part on at least a dynamic measure of capacity, wherein the requested object is served to the client from a server site other than the edge server site.

39. (Previously Presented) The computer program product of claim 38, the method further comprising:

if an object on the edge server site is no longer popular, as determined based on said dynamic measure of popularity of said object, deleting at least some part of the object from the edge server site.

40. (Previously Presented) The computer program product of claim 38, the method further comprising:

if an object on the parent server site is no longer popular as determined based on said dynamic measure of popularity of said object, and if the object is available on an origin server site, deleting at least some part of the object from the parent server site.

41. (Canceled)

42. (Previously Presented) The computer program product of claim 38, wherein replicating the portion of the requested object on said edge server site comprises:

replicating the portion of the requested object when said dynamic measure of popularity of the requested object is greater than a the dynamic replication threshold and there is enough storage on said edge server site to replicate the portion of the requested object;

otherwise, if there is not enough storage on said edge server site to replicate the portion of the requested object, then

i) comparing the popularity of the requested object, as determined based on said dynamic measure of popularity, against a dynamic measure of popularity of a least popular object in storage on the edge server site,

ii) if the dynamic measure of popularity of the requested object exceeds the dynamic measure of popularity of the least popular object in the storage, deleting at least some part of the least popular object from the storage, and

iii) repeating i) and ii) until enough storage is available for the portion of the requested object or until the dynamic measure of popularity of the requested object is less than the dynamic measure of popularity of the least popular object in the storage, and then

iv) replicating the portion of the requested object on the edge server if there is enough storage.

43. (Original) The computer program product of claim 38, wherein whether the requested object is popular is determined using at least a request rate for the requested object.

44. (Previously Presented) The computer program product of claim 38, wherein serving the requested object is performed separately from replicating the portion of the requested object.

45. (Previously Presented) A system for managed object replication and delivery, comprising:

a plurality of edge server sites in a network; and

a plurality of parent server sites in the network, said parent server sites distinct from said edge server sites,

wherein:

a request by a client for an object is directed to an edge server site in the network, and

if the edge server site has the requested object, the requested object is served to the client from the edge server site, and

otherwise, if the edge server site does not have the requested object, the client request is redirected to a first server site distinct from the edge server site, and,

if the edge server site does not have the requested object and the object is served to the client from a server site other than the edge server site, then a portion of the requested object is conditionally replicated on the edge server site for use in serving future client requests, said replicating being based at least in part on a dynamic measure of popularity of the requested object relative to a dynamic replication threshold, said dynamic replication threshold being based, at least in part, on at least one dynamic measure of capacity.

46. (Previously Presented) The system of claim 45, wherein redirecting the client request to a server site comprises redirecting the client request to a parent server site in the network and attempting to serve the requested object to the client from the parent server site.

47. (Previously Presented) The system of claim 45, wherein redirecting the client request to a server site comprises recursively redirecting the request until a parent server site in the network having the requested object is reached and serving the requested object to the client from the parent server site.

48. (canceled)

49. (Previously Presented) The system of claim 45, wherein directing a request by a client for an object to an edge server site comprises directing the request by the client for an object to a best or optimal edge server site.

50. (Previously Presented) The system of claim 49, wherein a best or optimal edge server site comprises an edge server site selected using at least one of a determination based on a best repeater selector, the likelihood of a copy of the requested object being available at the edge server site, and the bandwidth between the edge server site and the client.

51. (Previously Presented) The system of claim 45, wherein replicating of a portion of the requested object to the edge server site comprises replicating the portion of the requested object to the edge server site from a parent server site.

52. (Previously Presented) The system of claim 45, wherein at least one of the plurality of edge server sites and the plurality of parent server sites further replicate the portion of the requested object from an origin server site if the requested object is popular, as determined based on a the dynamic measure of popularity, and if the requested object is unavailable on parent server sites in the network.

53. (Previously Presented) The system of claim 45, wherein replicating the portion of the requested object to the edge server site comprises, if the requested object is unavailable on parent server sites in the network, replicating the portion of the requested object to the edge server site from an origin server site.

54. (Previously Presented) The system of claim 45, wherein said dynamic measure of popularity of the requested object is determined using at least a request rate for the requested object.

55. (Previously Presented) The system of claim 45, wherein at least one of the plurality of edge server sites and the plurality of parent server sites deletes at least some part of an object if the object is no longer popular, as determined based on said dynamic measure of popularity of the requested object.

56. (Canceled)

57. (Previously Presented) The system of claim 45, wherein replicating the portion of the requested object comprises:

replicating the portion of the requested object when said dynamic measure of popularity of the requested object is greater than a the dynamic replication threshold and

when there is enough storage on said the edge server site to replicate the portion of the requested object;

otherwise, if there is not enough storage on the edge server site to replicate the portion of the requested object, then

i) comparing the popularity of the requested object, as determined based on said dynamic measure of popularity of said object, against a dynamic measure of popularity of a least popular object in storage on the edge server site,

ii) if the dynamic measure of popularity of the requested object exceeds the dynamic measure of popularity of the least popular object in the storage, deleting at least some part of the least popular object from the storage, and

iii) repeating i) and ii) until enough storage is available for the portion of the requested object or until the dynamic measure of popularity of the requested object is less than the dynamic measure of popularity of a least popular object in the storage, and then

iv) replicating the portion of the requested object on said edge server site if there is enough storage.

58. (Previously Presented) The system of claim 45, wherein serving the requested object is performed separately from replicating the portion of the requested object.

59-65. (Canceled).

66. (Previously Presented) The method of claim 1 wherein the server site from which the requested object is served to the client is a peer server site of the particular edge server site.



67. (Previously Presented) The method of claim 1 wherein the server site from which the requested object is served to the client is the first server site.

68. (Previously Presented) The method of claim 1 wherein the server site from which the requested object is served to the client is a peer of the first server site.

69. (Previously Presented) The method of claim 1 wherein the step of conditionally replicating the portion of the requested object on the particular edge server site replicates the portion of the requested object from a peer server site of the particular edge server site.

70. (Previously Presented) The method of claim 1 wherein the step of conditionally replicating the portion of the requested object on the particular edge server site replicates the portion of the requested object from a server site.

71. (Previously Presented) The method of claim 1 wherein the dynamic measure of popularity of the requested object is based at least in part on one or more of:

- (a) a local dynamic measure of popularity of the object; and
- (b) information regarding the popularity of the object on other servers.

72. (Previously Presented) The method of claim 1 wherein the requested object comprises chunks, including initial chunks and remaining chunks, and wherein the portion of the requested object comprises only initial chunks of the object.

73. (Previously Presented) The method of claim 27 wherein the remaining chunks of the object are selectively replicated based at least in part on how many of the initial chunks are served to the client.

74. (Previously Presented) The method of claim 16 wherein the dynamic measure of popularity of the requested object is based at least in part on one or more of:

- (a) a local dynamic measure of popularity of the object; and
- (b) information regarding the popularity of the object on other servers.

75. (Previously Presented) The method of claim 16 wherein the requested object comprises chunks, including initial chunks and remaining chunks, and wherein the portion of the requested object comprises only initial chunks of the object.

76. (Previously Presented) The computer program product of claim 23 wherein the dynamic measure of popularity of the requested object is based at least in part on one or more of:

- (a) a local dynamic measure of popularity of the object; and
- (b) information regarding the popularity of the object on other servers.

77. (Previously Presented) The computer program product of claim 23 wherein the requested object comprises chunks, including initial chunks and remaining chunks, and wherein the portion of the requested object comprises only initial chunks of the object.

78. (Previously Presented) The computer program product of claim 38 wherein the dynamic measure of popularity of the requested object is based at least in part on one or more of:

- (a) a local dynamic measure of popularity of the object; and
- (b) information regarding the popularity of the object on other servers.

79. (Previously Presented) The computer program product of claim 38 wherein the requested object comprises chunks, including initial chunks and remaining chunks, and wherein the portion of the requested object comprises only initial chunks of the object.

80. (Previously Presented) The system of claim 45 wherein the dynamic measure of popularity of the requested object is based at least in part on one or more of:

- (a) a local dynamic measure of popularity of the object; and
- (b) information regarding the popularity of the object on other servers.

81. (Previously Presented) The system of claim 45 wherein the requested object comprises chunks, including initial chunks and remaining chunks, and wherein the portion of the requested object comprises only initial chunks of the object.

82. (Previously Presented) A method as recited in claim 1, wherein the at least one dynamic measure of capacity represents available capacity on the particular edge server.

83. (Previously Presented) A method as recited in claim 16, wherein the at least one dynamic measure of capacity represents available capacity on the edge server site.

84. (Previously Presented) A system as recited in claim 45, wherein the at least one dynamic measure of capacity represents available capacity on the edge server site.